Moving-Correlation-Code Triangulation Range Imaging

Frank Pipitone and Ralph Hartley
Navy Center for Artificial Intelligence
Naval Research Laboratory
Washington, DC 20375-5337

ABSTRACT

A new method is described for obtaining accurate range images at high speed in a low-cost instrument. A prototype has been built and tested, and a patent application submitted. The method resembles grid-coding in that a camera and a stripe projector are directed at a scene, but the projector is different. It consists of a thin light source (xenon tube and slit) on the axis of a turntable, and a binary mask conforming to a cylinder coaxial with this. The mask has alternate black and clear stripes parallel to the axis. It forms a DeBruijn sequence , i.e., a sequence in which all possible sub-sequences of given length n occur. No lens is used, deliberately smoothing the resulting illumination. In operation, the turntable rotates, and six consecutive images are taken at uniform intervals. A given pixel records six consecutive samples of a scene point. This six-vector, when normalized to unity to accommodate reflectance variations, is unique to the place in the sequence from which it came. Thus we can compute the position in 3-space of the surface point at which the pixel is looking. Observed accuracy is .1 millimeter at 30 centimeters range.

Keywords: range, rangefinder, triangulation, structured light.

1. INTRODUCTION

In recent decades, a wide variety of instruments have been built to obtain range images. A survey[1] by Besl describes many of the design options for such systems. A range image is a two-dimensional array of numbers which give the depth of a scene along many directions from the center of the instrument. That is, instead of measuring the brightness of many points in a scene, as in a TV camera, it essentially measures the location of each point in three-dimensional space. These instruments are divided into two broad categories, triangulation and time-of-flight. Our instrument is in the first class. Measures of their performance include range accuracy, pixel rate, frame rate, motion tolerance, angular field of view, minimum and maximum measurable range, and angular pixel separation. There are tradeoffs among these in choosing a system design. For example, laser point-scanning systems[2] can achieve very high accuracy, typically at the expense of speed. One of our earlier systems[3], a laser point scanner, achieves an extremely large field at the expense of both accuracy and speed. Structured light triangulation systems, such as grid-coding (discussed below) and our present method use parallelism to achieve high speed and with appropriate design, retain high accuracy.

We have conceived and prototyped a new range imaging sensor consisting of a camera and a structured-light projector. The structured-light projector produces a non-uniform pattern of stripes which is moved across the scene at a nominally constant angular velocity, during which several frames are acquired by the camera. This allows the reconstruction of one frame of range data. This work is intended to provide a low-cost, high performance instrument for obtaining range images suitable for various computer vision applications, such as the recognition and localization of objects, automatic inspection, and the recovery of computer models of surface shape from physical objects.

It is instructive to compare the present method with the existing method called grid-coding. Sato and Otsuki[4] describe one relatively recent version of this type and reference others. In this method, a scene is observed by a CCD camera and illuminated by a light source projected through an array of parallel opaque and transparent stripes of varying width. This pattern is usually formed by a programmable liquid crystal mask. A sequence of these patterns (e.g., 10) is projected, taking one CCD camera image per pattern. The patterns are constructed so that the sequence of pixel values (each 1 or 0) for one pixel in consecutive frames forms a grey code uniquely indicating the column of the projected array at which the pixel is looking. The grid-coding method requires the expense of a switchable mask, such as a liquid crystal mask. It has limited range accuracy; if the mask has N columns, only N range values can be discriminated, although in some systems columns of fine dots or thin stripes are projected within the mask columns, allowing a suitably high-resolution camera to measure their position to high (sub-pixel) accuracy. Lens optics are required in the projection system. Also, it requires at least log2(N) sequential patterns to obtain one range image. Our method provides an improvement over the grid coding approach in cost, range accuracy, depth of field, and time required to acquire a range image.

2. DESCRIPTION AND OPERATION

2.1 The general concept

We will now briefly describe this new range instrument in its most general form (see Fig. 1). One element is a detector consisting of one or more photosensitive elements, each with a narrow ray along which it is sensitive to light. For example, one could use a photodiode with a pinhole in front of it, with a lens focusing on the pinhole. Another reasonable variation is a CCD camera (as in the prototype). The detector's output is digitized periodically in time and stored in the memory of a computer. The detector is directed at a scene for which range data is desired.

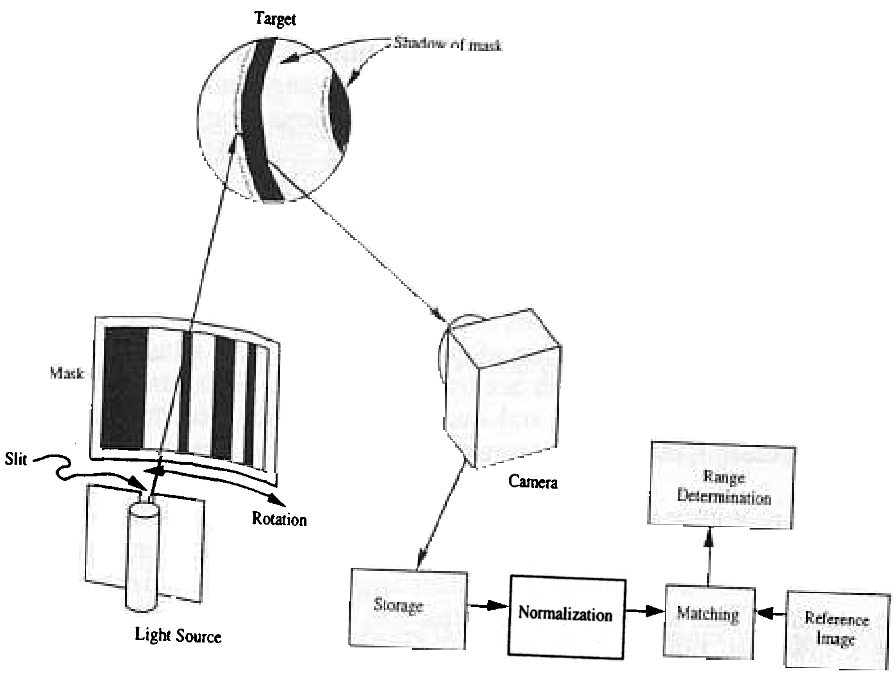

Figure 1. Schematic Illustration of System

A second element is a special structured-light projector. This projector is centered at a distance d from the detector, and directed so that it illuminates approximately the same region of the scene that is observed by the detector. The projector periodically emits a flash of structured light, each flash accompanied by digitization of the detector output. Each flash of structured light has the following description: The light emanates from a source (ideally cylindrical) on the central axis of the projector, an axis perpendicular to the baseline from detector center to projector center. The illumination is an analog stripe pattern, with the stripes corresponding to planes in space containing the projector's axis. The stripe pattern is therefore describable as a function of angle around the axis; $f(\theta)$. A key requirement is that the code have the property that a few regularly spaced samples of any region of the code form a unique vector, so that we can tell what region of the code they came from. Such patterns can be generated and projected over a wide range of distances and angles by analog smoothing of a binary pattern. This can be done without any lenses by illuminating the binary mask with a light source shaped like a thick line. Finally, to produce range data, a computer program generates one range point for each photodetector element. It does this by determining which place in the code the sequence of samples from a given photodetector element matches. Range is a monotonic function of this position. Essentially, range is computed by determining which plane containing the axis of the projector is being observed by a given photodetecting element. Then the line of sight of that element intersects that plane at a unique place, the location of the desired range sample.

2.2 The prototype system

As illustrated in Figure 1, our prototype system consists of a camera interfaced to a computer so that intensity images of a scene are recorded at appropriate times, and a projector. The system uses a 512X512 Dalsa camera with a 12.5 mm lens. The computer is a Hyperspeed i860 card hosted by a 486 PC. Figure 2 shows a photograph of the prototype system. The projector consists of a thin light source (a xenon flash tube with a slit in front) on the axis of a turntable, and a binary mask conforming to a cylinder coaxial with this axis. The mask has alternate black and clear stripes parallel to the axis.

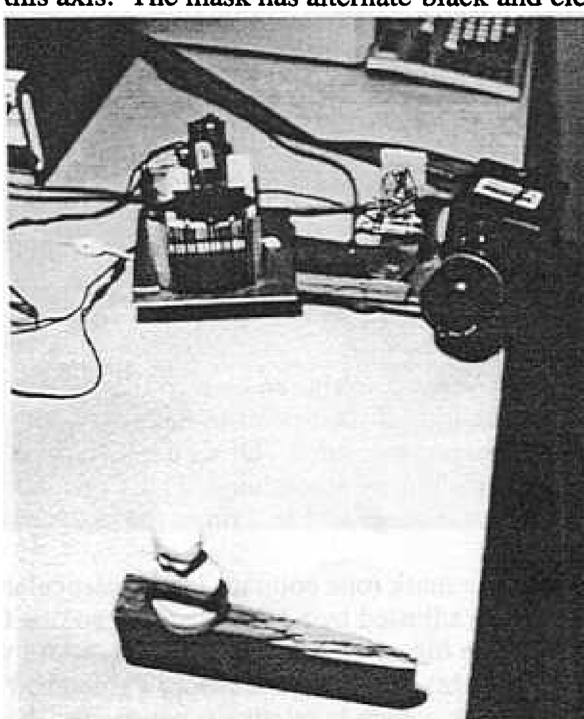

Figure 2. Photograph of the Prototype

The code on the mask forms a DeBruijn sequence[5], i.e., a sequence in which all possible sub-sequences of a given length n occur. For example, a code for n = 6 has $2^6 = 64$ columns such that any consecutive sample of 6 columns is unique. The following is the code used in our prototype:

111111000001000011000101001111010001110010010110111011001101010,

where 1 represents opacity and 0 represents transparency. The code 000000 was omitted because it may give poor position accuracy at its location. Figure 3 shows a mask formed by such a code. Four cycles are shown, although approximately three cycles are on the prototype. It should be used so that any given pixel of the camera observes a surface point illuminated only by stripes from one cycle of the code, to avoid range ambiguity. For example, we could use only a 1 or 2 cycle mask, restrict the location of the subjects in the scene, or position the camera so that it cannot see through a full cycle of the projected stripe pattern with any individual pixel.

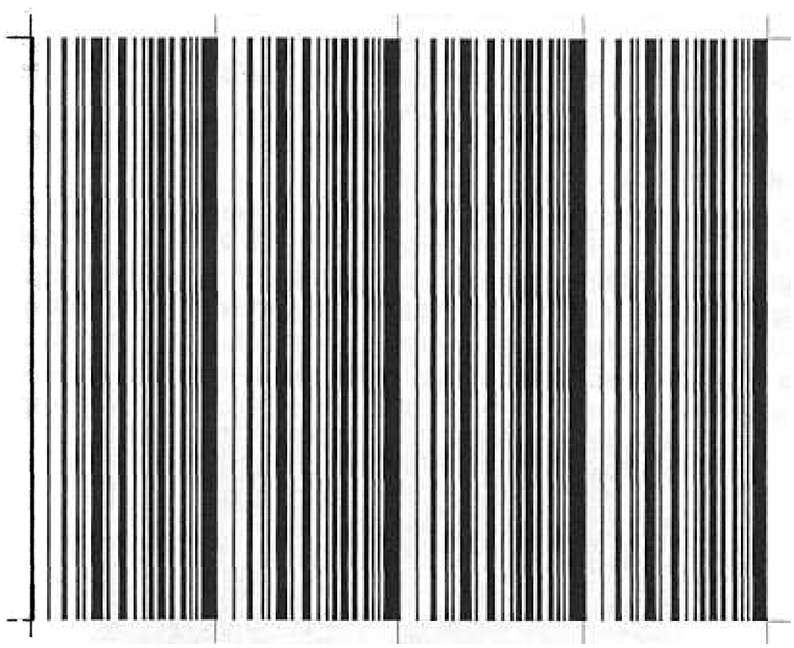

Figure 3. The Mask Used in the Prototype

The particular structured light pattern used was chosen by testing in computer simulation the ambiguity properties of candidate patterns. The most distant point in the pattern that matched (to a fixed level of accuracy) a given point in the pattern was measured. Of all the patterns tested the best was a DeBruijn sequence converted into an analog pattern by smoothing. The smoothing function was not critical as long as the smoothing interval was in the range of 1 to 2 times the width of a bit.

The width of the thinnest stripes in the mask (one column) is in a particular relationship with the diameter of the light source (which may be adjusted by a slit if needed) so that the shadow of the mask has a particular amount of blur (light source diameter 1 to 2 times stripe width works best). Thus, instead of projecting a sharp binary (dark and light) pattern on the scene, a smoothed version of this pattern is projected, without using a greyscale mask, which is relatively expensive. We used an ordinary computer printer and viewgraph transparency in our prototype. We used no lens in the projector. The depth of field is therefore very large, and no focusing is needed. The system can be implemented at various scales by changing the baseline distance between camera and projector. This is currently approximately 275 mm. Other geometric design parameters include the camera and lens parameters, the distance between light source and mask (now 95 mm), the code column width (now .5 mm) and the order n of the code.

In operation, the turntable is rotated approximately uniformly by a rotating cam resting against an arm attached to the turntable, and an image is taken, illuminated by a xenon flash, at uniform angular intervals. The interval is the angular width of one code column. This is done for six consecutive frames. Consider one pixel of the camera. In the six consecutive frames, that pixel records samples of the intensity of the projected light, as reflected from the small spot on the scene observed by the pixel. The six-vector consisting of those samples, when normalized so that its magnitude is always 1, to accommodate reflectance variations, is unique to the place in the sequence from which it came. Thus we can compute the position in 3-space of the surface spot at which the pixel is looking. This is possible because the pixel's ray defines a line in space trough the observed surface spot, and the place in the code sequence matching the observed 6-vector defines a plane through the observed surface spot (and through the projector center). Thus one can find the spot location as the intersection of that line and that plane. A computer program determines where in the sequence the observed 6-vector came from, thus determining the plane in space through the projector axis and through the observed spot in the scene for each pixel. Figure 4 shows a shaded rendering of a range image produced by the prototype system. The statue is 11 cm in height. The uniform region to its left is in the shadow of the structured light; range data is not available there.

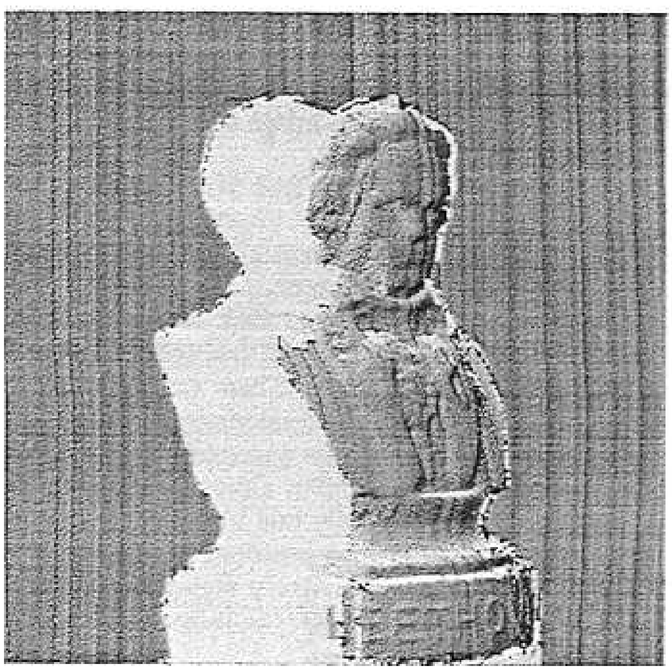

Figure 4. Shaded Rendering of a Range Image

3. COMPUTATIONAL AND CALIBRATION CONSIDERATIONS

Prior to imaging a desired scene, a set of six frames is taken of a reference plane, consisting of a flat white surface approximately 35 cm from the camera and perpendicular to its optic axis. These six reference images are taken as described above; at uniform intervals of the projector's motion, just as during range imaging of a desired scene. To obtain a range of a scene, a new set of six image frames, which we will call data images, is taken. Then for each pixel (i,j) of the data images, the best match is found with the corresponding row of the reference images. That is, k is found such that the six-vector at pixel (k,j) of the reference images is as close as possible, in Euclidean distance, to the six-vector at pixel (i,j) of the data images. The matching is not done by exhaustive search on k. First, the match result for neighboring pixel (i-1,j) is tried, exploiting the fact that many surfaces are fairly smooth. If this k value corresponds to a poor match, then a hash table scheme is used to index from the measured six-vector at pixel (i,j) to candidates for the best k value. In either case, a final local search on k is made, to optimize the match. Real valued k is obtained by interpolation, for greater accuracy.

The parameter k is essentially range in a peculiar coordinate system; it is the horizontal coordinate of the position on the reference plane at which the light source projects through the object point observed by pixel (i,j). We are currently developing a rectification procedure which will generate Cartesian range coordinates at each pixel.

Since xenon flashtubes frequently vary in discharge energy by an amount on the order of 10%, despite voltage regulation, we calibrated out the effects of this variation. We did this by exploiting a subset of pixels, which we call the calibration pixels, for which the reference plane is visible in both the reference images and the data images, which contain an object only partially occluding the reference plane. For each of the six projector positions, We divided the average value of the calibration pixels in the corresponding data image by the average value of the calibration pixels in the corresponding reference image. Then every pixel of the corresponding data image was divided by this correction factor.

4. EXPERIMENTAL RESULTS

The prototype system was adjusted so that the six-vectors produced by surface points at various places in the angular field of the projector were well separated. These adjustments included centering the slit around the rotation axis of the projector, aligning the mask stripes parallel to the slit, and adjusting the slit width to approximately 1 mm. Also, the timing of the xenon flashes was adjusted so that the six image frames were taken one mask column apart. Then various range images were generated and evaluated. The image of Figure 4 is a Lambertian rendering of a range image of an off-white statue of Beethoven, 11 cm high. The RMS range noise was approximately .09 mm, measured at the reference plane. The error on the statue itself appears similar, although we have not directly measured it. The vertical stripes in the rendered image are primarily due to variation in range variance as a function of code position. This is due to the fact that at some positions the six samples lie on points of the analog code waveform which have low slope, causing relatively high range variance, and for other positions at least some of the six samples lie on high-slope points, resulting in lower range variance.

Although this method works with only six frames of camera data per range image, we tried averaging several frames to reduce certain types of noise. Using N frames instead of one in each of the 6 positions appears to result in roughly the expected $1/\sqrt{(N)}$ reduction of range noise. For example, For N = 10, the RMS range noise was measured as .037 mm, versus .09 mm for one frame. The contributions to noise are primarily imperfect compensation for flash energy variation, non-repeatability of the cam mechanism and quantization and other noise in the camera's intensity values.

5. CONCLUSIONS AND FUTURE DIRECTIONS

We have built, tested, and submitted for patent a triangulation range imaging sensor based on a moving structured-light code pattern. A range accuracy of approximately .1 mm was achieved at a standoff of about 1 foot. Only six frames of camera data were needed to form a range image. The cost of the prototype was extremely low; a viewgraph transparency was used for the mask, and hobby-grade Xenon flash hardware was used. The mechanical components were low-precision. The main cost was for the camera, frame-grabber, and computer, which are standard off-the-shelf components. The system is easily reconfigured to different angular fields and range intervals. We expect to refine the system in various ways, including optimizing the design parameters described earlier and improving the quality of the mask and the repeatability of the rotational motion. Also, we will consider the merits of various design alternatives such as alternative rotation mechanisms and the use of a complete cylindrical drum for the code mask.

REFERENCES:

[1] Besl, P.,"Active, Optical Range Imaging Sensors", in Machine Vision and Applications, Springer-Verlag, Vol. 1, pp 127-152, 1988.

[2] Rioux, M., Blais, F., Beraldin, J., and Boulanger, P., "Range Imaging Sensors Development at NRC Laboratories", Proc. of the Workshop on Interpretation of 3-D Scenes, pp 154-160, Nov. 27, 1989, Austin, TX, IEEE Press.

[3] Pipitone, F., and Marshall, T., "A Wide-field Triangulation Laser Rangefinder for Machine Vision" International Journal of Robotics Research, Vol. 2, No. 1, spring 1983. (Also in SPIE Milestone Series Vol. 115, 1995)

[4] Sato, Y., and Otsuki, M., "Three-Dimensional Shape Reconstruction by Active rangefinder", Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp142-147, June, 1993.

[5] Jansen, C.J.A.; Franx, W.G.; Boekee, D.E., "An efficient algorithm for the generation of DeBruijn cycles", IEEE Transactions on Information Theory, vol.37, no. 5, pp 1475-8, September 1991.